

Connecting Researchers with Companies for University-Industry Collaboration

Qi Wang
School of Management,
Xi'an Jiaotong University
wangqijlu@sina.com

Weiwei Deng
Department of Information
Systems,
City University of Hong Kong
deng_weiwei@126.com

Jian Ma
Department of Information
Systems,
City University of Hong Kong
isjian@cityu.edu.hk

Mingyu Zhang
School of Management,
Putian University
simileyu@mail.ustc.edu.cn

Xiuwu Liao
School of Management,
Xi'an Jiaotong University
liaoxiuwu@xjtu.edu.cn

Abstract

Nowadays, companies are spending more time and money to enhance their innovation ability to respond to the increasing market competition. The pressure makes companies seek help from external knowledge, especially those from academia. Unfortunately, there is a gap between knowledge seekers (companies) and suppliers (researchers) due to the scattered and asymmetric information. To facilitate shared economy, various platforms are designed to connect the two parties. In this context, we design a researcher recommendation system to promote their collaboration (e.g. patent license, collaborative research, contract research and consultancy) based on a research social network with complete information about both researchers and companies. In the recommendation system, we evaluate researchers from three aspects, including expertise relevance, quality and trustworthiness. The experiment result shows that our system performs well in recommending suitable researchers for companies. The recommendation system has been implemented on an innovation platform, InnoCity.

1. Introduction

With the increasingly keen market competition, companies have been aware of the necessity and benefits of collaborating with other entities to develop new technology and products [4, 36]. Especially, companies prefer cutting-edge knowledge from universities and research institutes for advancing

innovations and improving company performance [4, 34]. Companies put lots of effort in building linkages with universities through multi-channels, such as technology licensing, collaborative research, contract research, and consultancy [12, 25]. Researchers are also motivated to engage with industry to gain reputation and sense of achievement from research commercialization as well as funds and insights to support their future research [8, 18]. However, scattered market players and high search cost hinder the connections between knowledge suppliers and seekers. Information of companies and researchers is distributed over thousands of disconnected websites and systems. Searching desired information is time consuming because of information asymmetry and disconnection. There is an urgent demand for a 'bridge' to connect researchers and companies, through which they can share and commercialize knowledge. It is beneficial for them since companies can obtain advanced expertise while researchers gain funds for further research. Several platforms have been established to meet such demands such as InnoCity, Yet2.com and InnoCentive. With more and more researchers disseminating knowledge on such platforms, they have become a large knowledge and expert repository for companies

With the development of platforms, information overload problem arises inevitably. There are too many experts on the platform for companies to discover desired ones. Therefore, we design a research social network based researcher recommendation system to facilitate the connections between companies and researchers. The research social network provides us with complete, creditable and

well-organized information about researchers and companies, aims to solve the scattered and asymmetric information problem. Our recommendation system recommends appropriate researchers according to companies' demand so that it can ease companies' search cost and help companies find experts in a certain domain.

Lots of efforts has been made to develop expert recommendation systems. The developed systems are applied in various contexts, such as expert finding in organizations [41], question answerer recommendation for online Q&A communities [37], co-author seeking in research social networks [32] and reviewer assignment for project selection [31]. However, few of them attempts to connect researchers with companies for knowledge commercialization and university-industry collaboration. Our work bridges this gap by designing a recommendation system to facilitate the connection between researchers and companies with the help of an online platform. We analyze the candidate researchers from three aspects: expertise relevance, quality and trustworthiness. We profile the expertise of researchers and companies by utilizing their authored publications and patents. Their expertise is then matched to decide whether they are in a similar domain. The quality is analyzed in terms of academic research, technology invention and project experiences. We build a weighted heterogeneous social network according to their previous collaboration activities. The social proximity of researchers to the target company is computed by a weighted Katz algorithm to represent the trustworthiness of researchers. Finally, we rank the candidates by integrating the analysis results of the above three aspects.

The rest of the paper is structured as follows. Section 2 reviews the related research. Section 3 introduces our researcher recommendation system in detail. In Section 4, we conduct a small-scale experiment to show the effectiveness of our recommendation algorithm. The last section concludes the contributions and implications of this research.

2. Related research

The recommendation system is a tool to provide suggestions for users which improves the efficiency of information searching. In this part, we generally review the main stream of recommendation systems and introduce the expertise profiling specifically.

2.1 Recommendation systems

Since mid-1990s, recommendation systems have attracted much attention. Recommendation systems are widely applied in online platforms [26]. Performance of the recommendation systems is the key for such platforms to satisfy their users and achieve success. Both academia and industry have put in much effort to design efficient algorithms and improve the performance of recommendation systems. There are three main streams of recommendation systems: content-based recommendation systems, collaborative filtering recommendation systems and hybrid recommendation systems [13, 17, 28, 33]. Content-based recommendation systems recommend items that are similar to user's previous preference. Content refers to the features of items, such as cinematic genre and music genre. User's preference is profiled based on the features of items they liked before. New items are recommended if their features match the user profile. Content-based recommendation systems have been criticized for overemphasizing the similarity while overlooking the diversity of recommended items. Collaborative filtering (CF) recommendation systems are split into user-based CF recommendation systems and item-based CF recommendation systems. User-based CF systems recommend items that are liked by like-minded users, who have interest in common items. And the systems recommend items that are liked by like-minded users but have not been discovered by the target user. Item-based systems recommend items similar to those liked by the target user. It disregards features or content of items and only uses the user-item relationship to analyze the item similarity. Items are similar if they are liked by the same group of users. Compared with content-based recommendation systems, CF recommendation systems can generate more diverse items. However, it encounters data sparsity and cold start problem. Data sparsity problem means that the user-item relationship matrix is too sparse to identify user preference and item similarity. Cold start problem refers to the problem that the systems cannot make recommendations for new users as the systems cannot identify their preference and new items cannot be recommended because they have few ratings to identify their similar items. Both content based systems and CF systems have advantages and drawbacks, so hybrid recommendation systems are developed to integrate the two kinds of strategies [11, 16, 38]. Nowadays, with the development of social network, recommendation systems tend to utilize the relationship between users contained in social network to make recommendations [13, 17]. To better identify similar items, semantic enhanced recommendation systems are also developed [29, 30]. In our research, we adopt a hybrid recommendation algorithm. We

analyze the interest domain based on the documents associated with the company. Profiles of researchers are also analyzed based on the content of authored documents. We also consider collaborative filtering strategy to recommend researchers who are socially in close proximity to the target company and trustable. This can be achieved by analyzing their social network.

2.2 Expert finding

Among various recommendation systems, expert finding/recommendation systems are one of the most important and needed systems [27]. This kind of systems is designed to identify a list of people who possess knowledge and expertise about a given topic/query. Expert finding systems are initially applied in organizations as knowledge bases to store the expertise and skills of employees for knowledge management [9]. Now it has been applied in 1) online question & answering community for problem answering [20, 23, 37], 2) research social network for collaborator seeking [6, 32, 39], 3) research management system for reviewer assignment [31] and 4) organizations for well-informed colleagues retrieval [27, 41].

The key component for expert recommendation system is to analyze expertise of users, which is to use a set of topics to describe their knowledge [2]. In traditional expert finding systems, employees are required to provide their skills and knowledge to profile themselves [9]. Chen et al. (2010) design an expert recommendation system through which domain experts from different organizations can share product empirical knowledge effectively to facilitate product knowledge consultation and thus enhance product market competitiveness [7]. Knowledge concepts declared by the experts are used to profile the expertise. Recently, more systems rely on diverse sources to obtain the expertise information. Documents associated with an individual are regarded as evidence for his expertise, such as blogs, posts and papers authored by him. Balog propose a method to model the expertise of an expert for expert finding based on his associated documents [1]. Wang et al. build expertise profile by merging all documents an expert authored previously [37]. In academic knowledge management field, the main sources to profile a researcher's expertise are his publications [6]. With the development of research social network, richer information are generated in the platform, such as project information and self-claimed interests. These information has also been used to profile a researcher [32, 39, 40].

Besides the expertise dimension, two additional factors (i.e. social relation [10, 14, 37, 39] and quality

[31, 32]) are widely considered in expert finding research. Social relation is used to measure the proximity of two entities in the network or to measure the position of a given entity in the network. Relations can be various in different contexts. For example, it can be the co-author relationship in article collaborator finding context [32]. Specially, quality criterion is widely adopted in academic context to represent the academic achievement of the researcher [31, 32]. This criterion is considered for assuring that the recommended researchers are excellent and have competence to complete the task.

One stream of research on university-industry collaboration aims to identify influence factors of the collaboration performance. There are some valuable findings for collaborator selection. De Fuentes and Dutrénit (2012) analyze the characters of researchers, which help foster collaboration via various channels and also bring long-term benefits for companies [12]. Factors such as seniority, academic status, previous collaboration experience, extent of collaboration, and research fields are identified for the collaboration. Perkmann, et al. (2013) systematically review the antecedents of academic researchers to engage with the industry [25]. The notable findings are that the previous collaboration experience with companies as well as quality and success in terms of scientific productivity and grant funding (from government or industry) is positively related to the engagement activities. Firms are more likely to collaborate with high-quality researchers [24]. Trust, personal relation and past experience influence the perceived success of collaboration and are usually taken into consideration when selecting partners [3].

Incorporating these valuable findings, we comprehensively analyze researchers from expertise relevance, quality and trustworthiness perspectives to identify suitable collaborators for companies. The theoretical foundations of the selected criteria are summarized in Table 1.

Table 1. Theoretical foundations of selected criteria

Expert finding research	
Relevance	Expertise profile [2, 31, 32]
Quality	Academic achievement [31, 32]
Trustworthiness	Social relation [10, 14, 37, 39]; Co-authorship [32]
University industry collaboration research	
Relevance	Research field [12]
Quality	Academic status [12]; Scientific productivity and grant funding [25]; High quality researchers [24]
Trustworthiness	Previous collaboration experience [12] [25]; Trust, personal relation and past experience [3]

3. Researcher recommendation system

Our recommendation system helps companies to find appropriate researchers and facilitate the connection between them, which may contribute to research commercialization. There are two modules in our system: the researcher analysis module and the recommendation module. The framework of our recommendation system is depicted in Figure 1. Needed data are collected based on a research social network, including demographic data about researchers and companies as well as publications, patents and projects owned by them. Our recommendation system can provide companies with decision support about university-industry collaboration, such as partner selection for collaborative research and consultancy. In the following sections, we introduce the researcher analysis module and the recommendation module in details.

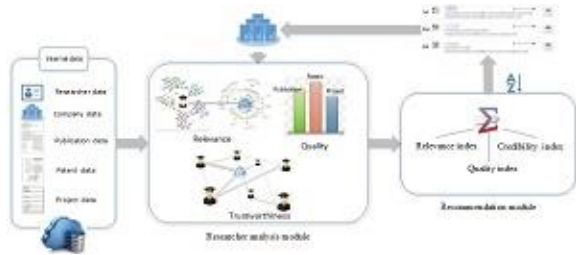


Figure 1. Framework of the system

3.1 Researcher analysis module

We analyze researchers from three aspects, including expertise relevance, quality and trustworthiness. The researcher analysis module consists of three parts. Each part analyzes one aspect for candidate evaluation. In relevance analysis part, we analyze the expertise of researchers and companies and then match the candidate researcher's profile with the target company's profile. In quality analysis part, we evaluate the candidate researcher's performance in academic achievement, technology invention and project experiences. In trustworthiness analysis part, trustworthiness is then measured by the social proximity between researchers and the target company in a heterogeneous collaboration network.

3.1.1 Relevance analysis. Relevance analysis evaluates whether a researcher has domain knowledge with regard to the company's demand. We first profile

the expertise of researchers and companies and then match their profiles for expertise relevance index.

Documents associated with entities, such as publications and patents, can be seen as evidence of their expertise and interest field. We collect publications and patents authored by researchers and companies to profile their expertise. To capture the main content of such documents, title, keywords, and abstract are selected to profile a document. Besides, we also extract the structured information to describe the content, such as the international patent classifications assigned to patents. We use information extraction techniques to process the text. Keywords extracted from free text are used to profile the document. Each document is processed as a bag of keywords merging the structured ontology classification and free text keywords. What is more, query information is collected if companies provide active query to discover researchers.

We use a semantic analysis method to compute the relevance index. Latent Dirichlet Allocation (LDA) is a widely used unsupervised learning algorithm for text modeling [5]. The basic assumption of this model is that a document is a mixture of various topics and each topic is represented by a set of words. This model can be used to analyze the similarity of documents and classify documents so that it can improve the performance of information retrieval [5]. We use this model to analyze the semantic similarity between documents and then to analyze the relevance between companies and researchers. The corpus used to train the model is the set of processed publications and patents.

We select Jensen Shannon divergence to measure the document similarity because each document is represented by a probability distribution. Jensen Shannon divergence is a popular method to measure the similarity between two probability distributions, which is based on Kullback-Leibler divergence [22]. The Jensen-Shannon divergence is defined as follows:

$$JSD(P_1, P_2) = \frac{1}{2} D(P_1, \frac{P_1 + P_2}{2}) + \frac{1}{2} D(P_2, \frac{P_1 + P_2}{2}) \quad (\text{Eq.1})$$

where $D(*, *)$ represents the Kullback-Leibler divergence defined by the following equation.

$$D(P, Q) = \sum_j p_j \ln \frac{p_j}{q_j} \quad (\text{Eq.2})$$

The relevance score of a researcher to a company is measured by the average similarity of authored documents.

$$rel_{ij} = \frac{\sum_{p \in doc_i} \sum_{q \in doc_j} JSD(p, q)}{|doc_i| |doc_j|} \quad (\text{Eq.3})$$

3.1.2 Quality analysis. Quality analysis evaluates the researcher's competence. In previous expert finding systems, quality of experts is ignored [7, 37, 41]. Researcher recommendation systems for co-author finding and reviewer assignment consider academic achievements (e.g. quality and quantity of publications) to measure the quality of researchers. However, performance in patents and projects is not less important than publication performance in our context. These indices are verified to be contributive to the performance of the commercialization activity [24, 25]. Therefore, we comprehensively consider the three aspects to evaluate a researcher's quality, including academic performance, patents and project experiences.

In the aspect of academic performance, we measure the quality of a researcher based on the quality and quantity of journal publications they authored. Referring to [31], we use the journal rank to represent the quality of publications. Journals are classified into three levels: level A, level B and level C. We measure the academic performance of researcher j as follows:

$$AP_j = \omega_A \cdot Q_{jA} + \omega_B \cdot Q_{jB} + \omega_C \cdot Q_{jC} \quad (\text{Eq.4})$$

Where ω_A , ω_B , and ω_C are the weights of different journal levels and $\omega_A > \omega_B > \omega_C$; Q_{jA} , Q_{jB} and Q_{jC} are the quantity of publications researcher j published in journals of corresponding level. In our research, we adopt a widely accepted journal ranking measurement, Journal Citation Reports (JCR) that is issued by the Intellectual Property and Science business of Thomson Reuters, to rank journals.

In the aspect of technology invention, we use the quality of patents invented by a researcher to evaluate the researcher's ability. Multiple indicators are identified to analyze patent quality, including forward citations to the patent, backward citations in the patent application, the number of independent claims, claim length, family size (i.e. the number of paralleled patents) and the number of patent applications in the previous year [19, 35]. Given the data availability and computation complexity, we select the most important indicators, forward citations and backward citations, as quality indicators. Then the performance of researcher j in technology invention is defined as follows:

$$IP_j = \sum_{i \in \text{patent}_j} fc_{ij} + bc_{ij} \quad (\text{Eq.5})$$

The project experience represents the ability of researchers to get fund from government and the industry. It is a positive signal for companies [25]. Project information cannot be obtained from a bibliographic database, but social networks are good sources for collecting such information. For example,

ScholarMate requires researchers to list their projects on their profile page. The number of project represents the researcher's quality in terms of acquiring public resources [25] (abbreviated as EP_i)

After scaling the three sub-indices into a unified interval, such as from zero to one, we aggregate the three sub-indices into productivity index.

$$qua_j = \alpha \cdot AP_j + \beta \cdot IP_j + \gamma \cdot EP_i \quad (\text{Eq.6})$$

where α , β and γ represent weights of the three sub-indices.

3.1.3 Trustworthiness analysis. Trust is an important factor when companies select collaboration partners due to the intrinsically high risk [3]. Trust between partners deeply influences the success of university-industry collaboration. Trustworthiness is defined as "the quality of a person or a thing that inspires reliability" in dictionary. In our research, trustworthiness reveals the character of a researcher, and represents whether the researcher can be trusted and easy to collaborate or not. Trustworthiness can be perceived based on previous collaboration experience, which means that companies have more confidence on researchers about their future collaboration if they have collaborated directly or indirectly before.

To quantify this perceptual and subjective character, we use the social proximity in the collaboration network to represent the degree to which the target company trusts the candidate researcher. The collaboration network is established based on multiple collaboration behavior (e.g. patent transfer, collaborations for publications, projects and patents). We build a weighted heterogeneous social network which is composed of researchers and companies. Edges of the constructed network represent the collaboration experiences, such as co-publishing papers, co-inventing patents, co-conducting projects, or transaction records of patents. Each edge has a weight representing the relation strength which is computed based on the collaboration frequency. In our system, the weight tw_{ij} of given edge is measured by aggregating the frequency of each kind of collaboration behavior between $entity_i$ and $entity_j$

Social proximity analysis has obtained extensive research. Mostly it is applied in link prediction. In other words, there is a greater probability to have a link between two nodes if they have high proximity score [21]. There are various approaches to analyze the proximity according to a certain social network. Liben-Nowell and Kleinberg conducted an experiment on large co-authorship network and compared the performance of nine common used approaches [21]. According to their findings, Katz's approach [15], a

classic path-ensemble based approach, performs better in co-authorship network. Therefore, we use Katz' approach to analyze the proximity between nodes in our heterogeneous collaboration network. The assumption of Katz's approach is that two nodes have stronger relationship if there are more paths between them and the length of those paths is shorter. If a certain researcher has a high proximity score to the target company, we can say that there is a higher likelihood that the company trust the researchers and is willing to collaborate with them.

The proximity matrix defined in [15] is represented as follows:

$$\mathbf{P} = \beta \cdot \mathbf{M} + \beta^2 \cdot \mathbf{M}^2 + \dots + \beta^k \cdot \mathbf{M}^k + \dots \quad (\text{Eq.7})$$

$$= (\mathbf{I} - \beta \mathbf{M})^{-1} - \mathbf{I}$$

where \mathbf{P} is the matrix of proximity score and \mathbf{M} is the adjacency matrix. The elements in the proximity matrix are measured as follows:

$$proximity_{ij} = \sum_{l=1}^{\infty} \beta^l \cdot |\text{path}_{i,j}^{(l)}| \quad (\text{Eq.8})$$

where $\text{path}_{i,j}^{(l)}$ is the set of all paths which start from i and end with j and are of length l . β ($0 < \beta < 1$) is a damping factor which means that the longer the path is, the lower importance the path has.

For weighted network, Liben-Nowell and Kleinberg set the $|\text{path}_{i,j}^{(l)}|$ as tw_{ij} [21]. It gives higher proximity scores for the direct and strong linked nodes. But for indirect links, weight information of relation is omitted to some extent. In our context, we think a company is more likely to collaborate with a researcher who has a strong relationship with its prior collaborators. The trust between two nodes is highly related to the trust degree between each pair of intermediate nodes. The relationship strength matrix \mathbf{S} is more suitable to be used to analyze the proximity value. We propose a weighted Katz algorithm and it is defined as follows:

$$\mathbf{P} = \beta \cdot \mathbf{S} + \beta^2 \cdot \mathbf{S}^2 + \dots + \beta^k \cdot \mathbf{S}^k + \dots \quad (\text{Eq.9})$$

$$= (\mathbf{I} - \beta \mathbf{S})^{-1} - \mathbf{I}$$

Most of the nodes are far away from the target company. It is meaningless to convince the company that the extremely remote 'relatives' are trustable to a low degree. Mathematically, from the equation for proximity computation, the long path has a low weight and it contributes little to the final score. Thus, we propose a pruning strategy to reduce the computation cost. We set n as a threshold of the shortest distance. If a node does not have any path to the company in n -step, we exclude it from the target company's network. It is an n -step iteration process. We initialize $C = \{\text{target company}\}$. While iteration time is less

than n , we expand set C by including neighbors of its members in each step, (i.e. $C = C \cup \{i \mid M_{ij} = 1 \wedge j \in C\}$). According to set C , we rebuild a smaller matrix \mathbf{M}' and its corresponding relationship strength matrix \mathbf{S}' for the proximity computation. Computation efficiency can be certified as the size of \mathbf{S}' is largely reduced. The proximity value of excluded nodes is set as zero without computation. Only length- n path is considered to measure the similarity. The trustworthiness value of researcher j for company i (abbreviated as tru_{ij}) equals to the corresponding element in matrix \mathbf{P} .

$$\mathbf{P} = \beta \cdot \mathbf{S}' + \beta^2 \cdot \mathbf{S}'^2 + \dots + \beta^n \cdot \mathbf{S}'^n.$$

3.2 Recommendation

In this module, we integrate the indices in all aspects to rank the candidate set. The ranking principle is that relevance is the basic and most important criterion. Quality and trustworthiness can be seen as extra bonus. In order to guarantee that all final recommended researchers are relevant to the company's domain, we adopt the cascade ranking process to refine the recommendation set and rank the candidates. There are two stages, one is pruning stage and the other stage is ranking stage.

In pruning stage, we use relevance criterion to refine the recommendation list, which is to exclude researchers with low relevance scores. We define the pruning function as $rel_{ij} < \delta$. δ is the lowest relevance threshold. It is a simple ranking strategy if $\delta=0$. Then we define a ranking function to rank the final recommendation list. The final score of each researcher can be represented as follows:

$$FS_j = rel_{ij} \times (1 + qua_j + tru_{ij}) \quad (\text{Eq.10})$$

Compared with quality and trustworthiness criteria, expertise relevance is the basic requirement for the recommended researcher. The researcher has a higher overall score if they possesses related domain knowledge. Besides, if they also performs well in this domain or has social relations with the target company, then they are ranked higher.

4. Experiment

We conduct an experiment to test the performance of our recommendation system. We select InnoCity (p0.innocity.com) as our experiment platform. InnoCity is an open innovation platform in China, and collects various innovation related resources (e.g. publications and patents), participators (e.g.

researchers, companies, and government departments) and opportunities (e.g. new projects and incentive policy). It has at least 5000 registered companies with detailed homepage information, more than 70,000 patents and more than 40,000 publications owned by companies. It connects to the largest research social platform in China (ScholarMate.com) that collects over 2 million researchers with their research output and interactions. Given its rich and complete information, we select this platform to collect data and evaluate our method.

We randomly select 30 companies from the platform. The researcher set is the full set of registered researchers. For each company, we generate twenty recommendations by different algorithms, including exact keyword match algorithm (baseline1), algorithm only with consideration of relevance dimension (baseline2), algorithm that considers relevance and quality dimension (baseline3) and our method. To test the effectiveness and advantages, we invite 15 experts to evaluate the recommendations for the selected companies. We select university professors as experts who have rich collaboration experience with industry companies. They are viewed as having the ability to judge whether the recommended researchers are suitable to collaborate with the target company or not.

We compute the relevance score, quality score and trustworthiness score of each researcher for each company to generate our recommendations. The process is described as follows:

In the first step, we use LDA model to analyze the similarity between documents of companies and that of researchers. We define the topic number as 100 and the corpus includes the titles, keywords and abstracts of patents and publications collected from InnoCity platform. After training, we use the topic probabilities and Jenson-Shannon diversion (Eq.1) to measure the similarity of each document pair. And the relevance score between companies and researchers is computed based on Eq.3.

In the second step, we measure the quality score for each researcher. Due to the data limitation, we only use the quantity factor to measure the performance of each aspect. We assign a same weight for publication performance, project performance and patent performance. The quality score is normalized by the maximum value.

In the third step, we build the collaboration network based on the author list in publications, patents and projects. The network contains 441524 edges and 63689 nodes. In our experiment, we consider the 2-step ego network for each company to save the computing cost. Utilizing the modified Katz's algorithm, we compute the trustworthiness score. We

compute the final ranking score using Eq.10 and select the Top 5 researchers as recommendation results.

Experts use 5-point Likert scale to show their subjective perception. Researchers recommended by four methods are mixed so that the assessment process is unbiased. We give brief information to introduce the recommended researcher in terms of demographic information, research achievements (i.e. publications, patents and reports) and project experience. Introduction of companies including their demand, information collected from homepage and patents information is given to assist experts' decision. To reduce subjective bias of experts, recommendations for each company are assessed by three experts. We compute inter-coder reliability of the three experts. The reliability should be larger than 0.7. The inconsistent evaluation will be eliminated. We average grades of consistent experts to assess each recommendation. Finally, our analysis set contains 24 companies sample and 1020 evaluation records.

The feedback of experts shows that our algorithm performs well. The average expert grade of recommendations generated by our algorithm is 3.51 and algorithm without trustworthiness has a score of 3.45 (Figure 2.). When only considering relevance factor, the average score is only 2.35. The average score of exact keyword match method is only 2.23. From Figure 3, we can see that the grades of recommendations generated by our algorithm and baseline 3 are always higher than that generated by baseline 1 and baseline 2. We also conduct paired-samples T test to verify whether there is significant difference between the four algorithms in expert grades. Results (line 2-5 in Table 2.) show our method and baseline 3 have significant higher grades than baseline1 and baseline 2 (95% confidence level). We can conclude that the quality of researchers affects the likelihood to collaborate with companies and this criterion is necessary for researcher analysis in researcher recommendation systems.

The difference between baseline1 and baseline 2 is not significant (the last line in Table 2.). The reason may be that the corpus used to train our LDA model can be enriched by adding more information of a documents and collecting more documents. We also find out that there is no significant difference between our method and baseline 3 (the first line of Table 2.). This might be because that 1) it is difficult to illustrate the indirect relation among companies and recommended researchers in our survey so that experts may not effectively perceive the trustworthiness information when evaluating the recommendations; 2) direct and indirect relations among researchers and companies are limited in our experiment so that the recommendation lists generated by our algorithm are

very similar to the lists generated by baseline 3. But from Figure 4, we can clearly see that the collaborators of the company ZTE almost have had collaboration experiences. In another word, the company tends to collaborate with the collaborators of its collaborators. It provides the evidence supporting our assumption that previous collaboration experience contributes to the trustworthiness and increases the likelihood of future collaboration. Although the improvement with trustworthiness is not significant in our experiment, it is still necessary to be considered as a feature to recommend researchers.

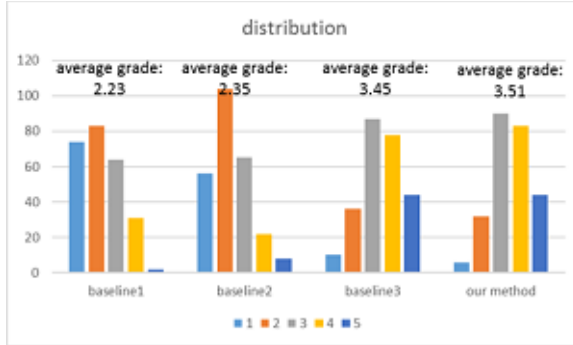


Figure 2. Score distributions



Figure 3. Average expert grade for each company

Table 2. Result of paired t test

Pair	Mean	t	Sig.
Our method - baseline3	.05714	1.403	.162
Our method - baseline2	1.12245	14.109	.000
Our method - baseline1	1.17623	14.120	.000
baseline3 - baseline2	1.06531	12.943	.000
baseline3 - baseline1	1.11885	13.269	.000
baseline2 - baseline1	.04508	.573	.567



Figure 4. Part of the collaboration network of company "ZTE". The larger orange circle represents the company ZTE and other circles are researchers who have collaborated with ZTE. The edge linked two circles represents they have collaborated before.

We also investigate experts which aspects they consider when conducting the assessment. Experts have consistent opinions that the recommended researchers must have knowledge related to the company. More than half of the experts emphasized the importance of qualifications and achievements of researchers. Some of them also think the trust relation may contribute to the collaboration, even though trust information is not effectively conveyed in our survey. Our method recommends researchers with higher quality (i.e. 0.4) and trustworthiness (i.e. 0.5) (Figure 5.). Recommendations generated by exact keyword match algorithm have the lowest average quality value and trustworthiness value.



Figure 5. Average quality and trust value

5. Conclusion

Various platforms are emerging to connect knowledge suppliers and knowledge seekers with the pressing need for knowledge sharing and transfer. Such platforms improve the knowledge utilization and benefit two parties. However, information overload has become a severe problem with the information

explosion. This paper proposes a recommendation system that aims to connect companies with desired researchers. It extends the general experts recommendation/finding systems, and combines characters of researchers and important factors to make it customized for university-industry collaboration. Relevance, quality and trustworthiness criteria are selected as rational and important criteria to assess researchers for further collaboration. The recommendation system is expected to facilitate the collaboration between companies and researchers, and further improves the knowledge transfer and research outreach.

Limitations and future work are also summarized. First, expertise relevance, quality and trustworthiness are considered in our system based on a thorough survey of previous research, but other factors such as personality issue may also have influence on potential collaboration. Future work will survey university-industry collaboration activities and summarize more significant factors. Second, the subjective evaluation will be improved by conducting longitudinal analysis to objectively record the future collaboration. Third, trustworthiness is an important criterion that needs more proofs. We plan to investigate the existing and future relationship among researchers and companies to enhance the criterion evaluation. Last, the proposed recommendation system tries to solve the information overload problem faced by two parties for further collaboration, but has minor effect when the two parties have engaged in collaboration. There will be lots of works to be done for university-industry collaboration.

Acknowledgement

The authors would like to thank the track chairs and reviewers for their comments. This work is supported by the National Natural Science Foundation of China (NO: 71546119, M1552003).

References

- [1] Balog, Azzopardi and de Rijke, Formal models for expert finding in enterprise corpora, (2006) 43.
- [2] Balog, Bogers, Azzopardi, De Rijke and Van Den Bosch, Broad expertise retrieval in sparse data environments, in: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, (ACM, 2007), pp. 551-558.
- [3] Barnes, Pashby and Gibbons, Effective University-Industry Interaction: A Multi-case Evaluation of Collaborative R&D Projects, European Management Journal, 20(3) (2002) 272-285.
- [4] Belderbos, Carree and Lokshin, Cooperative R&D and firm performance, Research Policy, 33(10) (2004) 1477-1492.
- [5] Blei, Ng and Jordan, Latent dirichlet allocation, the Journal of machine Learning research, 3(2003) 993-1022.
- [6] Chaiwanarom and Lursinsap, Collaborator recommendation in interdisciplinary computer science using degrees of collaborative forces, temporal evolution of research interest, and comparative seniority status, Knowledge-Based Systems, 75(2015) 161-172.
- [7] Chen, Chen and Wu, An expert recommendation system for product empirical knowledge consultation, in: Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, (IEEE, 2010), pp. 23-27.
- [8] D'Este and Perkmann, Why do academics engage with industry? The entrepreneurial university and individual motivations, The Journal of Technology Transfer, 36(3) (2010) 316-339.
- [9] Davenport and Prusak, Working knowledge: How organizations manage what they know, (Harvard Business Press, 1998).
- [10] Davoodi, Afsharchi and Kianmehr, A social network-based approach to expert recommendation system, in: Hybrid Artificial Intelligent Systems, (Springer, 2012), pp. 91-102.
- [11] De Campos, Fernández-Luna, Huete and Rueda-Morales, Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks, International Journal of Approximate Reasoning, 51(7) (2010) 785-799.
- [12] De Fuentes and Dutrénit, Best channels of academia-industry interaction for long-term benefit, Research Policy, 41(9) (2012) 1666-1682.
- [13] Debnath, Ganguly and Mitra, Feature weighting in content based recommendation system using social network analysis, in: Proceedings of the 17th international conference on World Wide Web, (ACM, 2008), pp. 1041-1042.
- [14] Fazel-Zarandi, Devlin, Huang and Contractor, Expert recommendation based on social drivers, social network analysis, and semantic data representation, in: Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, (ACM, 2011), pp. 41-48.
- [15] Katz, A new status index derived from sociometric analysis, Psychometrika, 18(1) (1953) 39-43.
- [16] Klačnja-Milićević, Vesin, Ivanović and Budimac, E-Learning personalization based on hybrid recommendation strategy and learning style identification, Computers & Education, 56(3) (2011)

885-899.

- [17] Konstas, Stathopoulos and Jose, On social networks and collaborative recommendation, in: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, (ACM, 2009), pp. 195-202.
- [18] Lam, What motivates academic scientists to engage in research commercialization: 'Gold', 'ribbon' or 'puzzle'?, *Research Policy*, 40(10) (2011) 1354-1368.
- [19] Lanjouw and Schankerman, Patent quality and research productivity: Measuring innovation with multiple indicators*, *The Economic Journal*, 114(495) (2004) 441-465.
- [20] Li, Liao and Lai, A social recommender mechanism for improving knowledge sharing in online forums, *Information Processing & Management*, 48(5) (2012) 978-994.
- [21] Liben - Nowell and Kleinberg, The link - prediction problem for social networks, *Journal of the American society for information science and technology*, 58(7) (2007) 1019-1031.
- [22] Lin, Divergence measures based on the Shannon entropy, *Information Theory, IEEE Transactions on*, 37(1) (1991) 145-151.
- [23] Liu, Chen and Huang, QA document recommendations for communities of question-answering websites, *Knowledge-Based Systems*, 57(2014) 146-160.
- [24] Perkmann, King and Pavelin, Engaging excellence? Effects of faculty quality on university engagement with industry, *Research Policy*, 40(4) (2011) 539-552.
- [25] Perkmann, Tartari, McKelvey, Autio, Broström, D'Este, Fini, Geuna, Grimaldi, Hughes, Kramel, Kitson, Llerena, Lissoni, Salter and Sobrero, Academic engagement and commercialisation: A review of the literature on university-industry relations, *Research Policy*, 42(2) (2013) 423-442.
- [26] Ricci, Rokach, Shapira and Kantor, *Recommender Systems handbook*, (2011).
- [27] Serdyukov, Taylor, Vinay, Richardson and White, Automatic people tagging for expertise profiling in the enterprise, in: *Advances in Information Retrieval*, (Springer, 2011), pp. 399-410.
- [28] Shambour and Lu, A trust-semantic fusion-based recommendation approach for e-business applications, *Decision Support Systems*, 54(1) (2012) 768-780.
- [29] Shi, Zhou, Kong, Yu, Liu and Wang, Heterocom: a semantic-based recommendation system in heterogeneous networks, in: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, (ACM, 2012), pp. 1552-1555.

- [30] Shishehchi, Banihashem and Zin, A proposed semantic recommendation system for e-learning: A rule and ontology based e-learning recommendation system, in: *Information Technology (ITSim), 2010 International Symposium in*, (IEEE, 2010), pp. 1-5.
- [31] Silva, Guo, Ma, Jiang and Chen, A social network-empowered research analytics framework for project selection, *Decision Support Systems*, 55(4) (2013) 957-968.
- [32] Sun, Xu, Ma and Sun, Leverage RAF to find domain experts on research social network services: A big data analytics methodology with MapReduce framework, *International Journal of Production Economics*, 165(2015) 185-193.
- [33] Symeonidis, Tiakas and Manolopoulos, Product recommendation and rating prediction based on multi-modal social networks, in: *Proceedings of the fifth ACM conference on Recommender systems*, (ACM, 2011), pp. 61-68.
- [34] Tödtling, Lehner and Kaufmann, Do different types of innovation rely on specific kinds of knowledge interactions?, *Technovation*, 29(1) (2009) 59-71.
- [35] Trappey, Trappey, Wu and Lin, A patent quality analysis for innovative technology and product development, *Advanced Engineering Informatics*, 26(1) (2012) 26-34.
- [36] Vega-Jurado, Gutierrez-Gracia and Fernandez-de-Lucio, Does external knowledge sourcing matter for innovation? Evidence from the Spanish manufacturing industry, *Industrial and Corporate Change*, 18(4) (2009) 637-670.
- [37] Wang, Jiao, Abrahams, Fan and Zhang, ExpertRank: A topic-aware expert finding algorithm for online knowledge communities, *Decision Support Systems*, 54(3) (2013) 1442-1451.
- [38] Wen, Fang and Guan, A hybrid approach for personalized recommendation of news on the Web, *Expert Systems with Applications*, 39(5) (2012) 5806-5814.
- [39] Xu, Guo, Hao, Ma, Lau and Xu, Combining social network and semantic concept analysis for personalized academic researcher recommendation, *Decision Support Systems*, 54(1) (2012) 564-573.
- [40] Yang, Ma, Silva, Liu and Hua, A Multilevel Information Mining Approach for Expert Recommendation in Online Scientific Communities, *The Computer Journal*, (2014) bxu033.
- [41] Yimam-Seid and Kobsa, Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach, *Journal of Organizational Computing and Electronic Commerce*, 13(1) (2003) 1-24.